DOCUMENT RESUME

ED 461 260 FL 024 540

AUTHOR Ehrman, Madeline
TITLE A Study of the Modern Language Aptitude Test for Predicting Learning Success and Advising Students.
PUB DATE 1994-09-00
NOTE 28p.; In: Language Aptitude Invitational Symposium Program Proceedings (Arlington, VA, September 25-27, 1994); see FL 024 538.
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Academic Advising; Comparative Analysis; *Individual Differences; *Language Aptitude; Language Tests; *Measurement Techniques; *Personality Traits; Prediction; Second Language Learning; Standardized Tests; Statistical Analysis; Test Use; *Testing
IDENTIFIERS *Modern Language Aptitude Test; Noncognitive Attributes

ABSTRACT

The Modern Language Aptitude Test (MLAT) was part of a project examining biographical, motivational, attitudinal, personality, and cognitive aptitude variables in 1,000 adult students preparing at the Foreign Service Institute for overseas assignments, with various subsamples completing different instruments. Data were analyzed using correlation, ANOVA, chi-square, and mulitple regression as appropriate to the data and research questions. The MLAT proved the best of the available predictors of language learning success. As part of an effort to expand the concept of language learning aptitude beyond the strictly cognitive, this study related the MLAT not only to end-of-training proficiency outcomes, but also to personality disposition, using both overall correlational data and information on extremely strong and weak learners. The MLAT has been found about equivalent in its current predictive power (based on correlation) to the time it was developed. In addition, it was found to be especially powerful at the extremes of performance as measured by speaking proficiency. Links between high MLAT scores and various other individual characteristics, including personality variables, emerged. Qualitative findings from use of MLAT part scores in student counseling also suggest utility for this well-established instrument beyond prediction of learning success. (Contains 45 references.) (Author/MSE)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 461 260

A STUDY OF THE MODERN LANGUAGE APTITUDE TEST FOR PREDICTING LEARNING SUCCESS AND ADVISING STUDENTS

by
Madeline Ehrman
Foreign Service Institute, U.S. Department of State

FL 024 540

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Madeline Ehrman
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

A Study of the Modern Language Aptitude Test For Predicting Learning Success and Advising Students

Madeline Ehrman
Foreign Service Institute, U.S. Department of State

The Modern Language Aptitude Test (MLAT) was part of a project examining biographical, motivational, attitudinal, personality, and cognitive aptitude variables among a total of 1,000 adult students preparing for overseas assignments at the Foreign Service Institute (with various smaller Ns for subsamples completing different instruments). Data were analyzed by correlation, ANOVA, chi-square, and multiple regression as appropriate to the data and the research questions. The MLAT proved the best of the available predictors of language learning success. As part of an effort to expand the concept of language learning aptitude beyond the strictly cognitive, this study relates the MLAT not only to end-of-training proficiency outcomes but also to personality dispositions, using both overall correlational data and information on extremely strong and weak learners. Qualitative findings from use of the MLAT part scores in student counseling activities are also described, suggesting utility for this well-established instrument beyond prediction of learning success.

This paper describes findings of research in progress at the Foreign Service Institute (FSI), a government language training institution. For years, incoming students have taken the Modern Language Aptitude Test (MLAT); indeed, a sample from FSI was among the groups on which the MLAT was originally normed (Carroll & Sapon, 1959). It is still in use as part of the agency's procedures for assignment to foreign language training. (Language aptitude testing is also done at other agencies.)

Over recent years the MLAT has become the subject of some controversy at FSI. Some program managers continue to see a good relationship between performance on the MLAT and in language training, others protest that the relation, such as it is, is not very strong and furthermore the MLAT may be not represent the true ability of those who lack formal education (Rockmaker, personal communication, 1993). Anti-MLAT opinion has also suggested that the MLAT was designed for the audio-lingual methodology that was in vogue in the late 1950's and 1960's and that the test is no longer valid for the much more "communicative" teaching that is now done at FSI (Bruhn, personal communication, 1992). Much of the distrust of the MLAT is doubtless connected with the increased suspicion of psychological testing during the last quarter century (Anastasi, 1988). The project on which this paper reports was initiated in order to take such questions about the MLAT out of the realm of allegation and find out just how useful it still is.

The present paper reports on two efforts to answer these questions. One is a quantitative investigation using a large sample of FSI students taken between 1992 and 1994. That study looks at the MLAT primarily as a predictor of language learning success in the FSI setting of intensive, full-time language learning for communicative use. The other portion of the paper describes a less rigorous attempt to make use of patterns of high and low MLAT part scores with individual students. The initial outcomes of this attempt, still highly exploratory, suggest that the

MLAT may have value for pinpointing areas of learning success and difficulty for a wide range of students, including some relatively able but context-dependent ones not well served by relatively grammar-oriented instruction.

Review of Literature

The MLAT was perhaps the culmination of a long tradition of psychometric test development and efforts to predict language learning achievement, and it achieved a fairly respectable level of success in the audio-lingual and grammar-translation classrooms of the 1950's and 1960's (Spolsky, 1995). Other important language aptitude tests developed out of the same tradition include the Pimsleur Language Aptitude Battery (Pimsleur, 1966) the Defense Language Aptitude Battery (Petersen & Al-Haik, 1976), and VORD (Parry & Child, 1990). The Pimsleur is different from the MLAT in particular because it includes a portion directly addressing the ability to infer language structure from an artificial language stimulus. The DLAB consists primarily of such induction-testing items, in a modified English. VORD was designed to test the ability to cope with the grammar of languages in the Altaic family and consists of items that test such grammatical prowess (Parry & Child, 1990). All four, including the MLAT, were found to have similar predictive validity (Parry & Child, 1990). This paper will not address these other instruments but will focus on the MLAT, which is the instrument that is still in use at the Department of State.[1]

The outcome of a major research project at Harvard University, the MLAT is based on a factor analysis of a large number of individual characteristics thought to contribute to language learning. Carroll (1962) describes the project in extensive detail; the MLAT Manual (Carroll & Sapon, 1959) provides information on the validation studies. The individual characteristics were grouped into four main categories: phonetic coding ability (distinguishing sounds and reflecting them graphically), grammatical sensitivity (recognizing and using syntactic relationships), memory (rote and contextualized), and inductive language learning. All but the last of these four are directly addressed in the five parts of the MLAT (see Figure 1).

Other components listed by scholars of language aptitude include motivation and knowledge of vocabulary in the native language (Pimsleur, 1968), the ability to hear under conditions of interference (Carroll, 1990), the ability to "handle decontextualizd language" (Skehan, 1991),and the ability to shift mental set and cope with the unfamiliar (Ehrman, 1994b, 1995, 1996; Ehrman & Oxford, 1995).

A desire for better prediction of language learning and the ability to exploit aptitude testing further has led to recent research efforts. At least two major projects in recent years have examined the role of individual differences in addition to strictly cognitive aptitude in language learning: the Defense Language Institute's Language Skill Change Project (Lett & O'Mara, 1990) and the Foreign Service Institute's Language Learning Profiles Project (Ehrman, 1993, 1994, 1995, 1996; Ehrman & Oxford, 1995; Oxford & Ehrman, 1995) investigated such variables as biographic factors, personality, motivation, anxiety, and learning strategies, as well general

[1] The remainder of the literature review owes much to a draft prepared by Frederick Jackson for an FSI roundtable at the Language Testing Research Colloquium in 1994 (Jackson, 1994).

intelligence (DLI only). A similar project has begun at the Central Intelligence Agency language school, though without personality variables, and DLI is engaged in a large-scale effort to improve the DLAB (Thain, 1992; Lett & Thain, 1994). This paper is part of one of these projects at FSI.[2]

Across a number of studies, predictive validity correlations for the MLAT have generally ranged between .42 and .62 for most languages, with outliers of .27 for certain non -Indo-European languages at the Defense Language Institute and as high as .73 for language instructor performance ratings at FSI (Carroll & Sapon, 1959). More recent tests of the MLAT are quite mixed. For instance, Brecht, Davidson and Ginsburg (1993) did not find the MLAT predictive of overall oral proficiency in intensive language training in Russia, though for the same programs they found Part III (Spelling Clues) to be "highly significant" in predicting listening comprehension and the Total Score to be significantly predictive of reading proficiency. They speculate that the lack of predictive value for oral proficiency is because this is a "communicative task." This suggestion is quite consistent with the questions raised at FSI (see above) and the point of view that standard aptitude measures do not "take into account" such developments as focus on communicative competence, pragmatics and discourse, new thinking by cognitive psychologists, etc. (Parry & Stansfield, 1990).

Another finding is that of Spolsky (1995), who reports that MLAT Part I correlated significantly with success on the part of Israeli learners of French as a foreign language, but the MLAT did not predict achievement in Hebrew at the same school, a variance he suggests may be related to differences in such factors as motivation, which is so powerful that it may override aptitude. (I suggest that it may also be the case that the students were learning Hebrew as a second language, not a foreign language, so not all their learning was classroom-based, which is the task for which existing language aptitude tests were designed.)

Most of the research cited addresses the use of the MLAT (and other aptitude measures) as predictors of learning success; and indeed this is an important consideration for assignment to intensive and long-term language training at taxpayer expense. However, a measure like the MLAT also has potential utility for *placement* in a program (Wesche, 1981) and *diagnosis* of learning difficulties, for counseling students, and for tailoring programs to their needs (e.g., Demuth & Smith, 1987; Sparks, Ganschow, & Patton, 1995). These applications have received far less attention in the literature. They are also among the areas of interest for the FSI investigation, and it is in these that the MLAT has been successfully used (Lefrancois & Sibiga, 1986; Wesche, 1981).

Methodology

Sample

In this study, there are 343 students altogether with at least a single Index score; of these, part scores for the subscales are available for 296. The mean age of the members of the sample is 37,

[2] The MLAT Project is separate but overlaps with the Language Learning Profiles Project, especially because for now it is using the same data set.

SD 9. Males constitute 59% and females 41% of the sample. The average age of students is 39, with a standard deviation of 9. The median education level is between bachelors and masters degrees. Of those that report previous language study, the average number of languages studied is 1.8. In the presentation of correlations with other instruments, *N*s are smaller, because not every person with an MLAT score in the data set completed all the other instruments.

FSI trains and tests students not only from its parent agency, but also from many other agencies. Student composition by agency and descriptions of student occupations in the sample at FSI would make identity of the institution obvious and is therefore omitted in this version.

Students in this study are beginners in long-term (i.e., 16 weeks or above) intensive language training. The languages they are studying are classified into four categories based on agency experience with the length of time needed by English speakers to reach "professional" proficiency (S-3 R-3—see 'Instrumentation' for a brief description of the ILR rating scale): 1. Western European; 2. Non-Western European but relatively quick for English speakers to learn (Swahili, Indonesian, and some North European languages); 3. Other non-Western European but excluding the category 4 languages (e.g., Russian, Thai); 4. "Super-hard" languages (Arabic, Chinese, Japanese, Korean)[3]. Usual training lengths vary by language category. Most FSI students are expected to reach "professional" proficiency (S-3 R-3) in 24 weeks in a category 1 language, in 32 weeks in a category 2 language, in 44 weeks in a category 3 language, and in 88 weeks (2 academic years) in a category 4 language.[4] These expectations are normally reflected in the lengths of student assignments to training and are also taken account of in the statistics reported in this paper.

Instrumentation

The MLAT. The Modern Language Aptitude Test (MLAT), (Carroll & Sapon, 1959) is the classic language aptitude test, with 146 items. The manual describes its five parts: I: number learning (memory, auditory alertness); II: phonetic script (association of sounds and symbols); III: spelling clues (English vocabulary, association of sounds and symbols); IV: words in sentences (grammatical structure in English); and V: paired associates (memorizing words), together with a total score. The MLAT was correlated .67 with the Primary Mental Abilities Test (Wesche, Edwards, & Wells, 1982), suggesting a strong general intelligence factor operating in the MLAT. Split-half reliabilities for the MLAT are .92–.97, depending on the grade or age. For college students, validity coefficients are .18–.69 for the long form of the MLAT and .21–.68 for the short form. For adult students in intensive language programs, validity coefficients are .27–.73 for the long form and .26–.69 for the short form (Carroll & Sapon, 1959). This study used the long form.

The subscales of the MLAT are described briefly in Figure 1. The Index Score used at FSI originated in the 1960's as a T-score based on the Total score, with three standard deviations of

---

[3] The Department of Defense uses a similar classification.

[4] Only three percent of students in this sample were studying category 2 languages—too small a number for most analyses. Category 2 and 3 languages are therefore combined for certain analyses.

10 on either side of a mean of 50.[5] It has since become frozen as a translation of the Total, much like Scholastic Aptitude Test ratings until recently, because of the agency personnel system's dependence on over 30 years of Index records. For users of the MLAT who are more familiar with the raw Total score, a table of equivalencies is provided in Appendix A.

Note that *Index 50 is the mean established when the MLAT was originally normed* and includes a variety of subjects from high schools and colleges. Whether it in fact is still representative of the population outside FSI is uncertain. What is certain, however, is that a mean Index of 50 is no longer valid for FSI students. There has been a gradual upward tendency in the MLAT Index mean at FSI over the intervening 30 years: Wilds (1965) reported a mean Index of 54 (N=957, no SD); an agency-internal document reports a 1984 mean Index of 59 , SD 10, N-312 (Adams, 1984); and the mean Index for all the students in the current sample who had MLAT scores is 63 SD 10, N = 343.[6]

Figure 1. MLAT Subscales

Part I. Number Learning: This subtest requires the examinee to learn four morphemes and interpret them in combinations that form numbers; it is entirely orally delivered. The subtest is described in the Manual (Carroll & Sapon, 1959) as measuring part of memory and "auditory alertness" which play a part in auditory comprehension (showing how well one understands what one hears) of a foreign language.

Part II. Phonetic Script: This subtest requires the examinee to select a written equivalent (in Trager-Smith phonemic transcription) for an orally delivered stimulus. The MLAT Manual describes the subtest as dealing with the ability to associate a sound with a particular symbol, as well as how well one can remember speech sounds. In addition, the subtest is described as tending to correlate with the ability to mimic speech sounds and sound combinations in a foreign language.

Part III. Spelling Clues: In this entirely written subtest, an English word is presented in a very non-standard spelling. The examinee must select the correct synonym. Vocabulary items are progressively more difficult, though the most difficult is probably within the repertoire of a college graduate. According to the Manual, scores on this part depend largely on how extensive a student's English vocabulary is. As in Part II, it measures the ability to make sound-symbol associations but to a lesser degree.

Part IV. Words in Sentences: The stimulus is a sentence with an error. The examinee must indicate which part of another sentence matches the designated part. The subtest is entirely in writing. It is described as dealing with the examinee's sensitivity to grammatical structure and thus expected to provide information about the ability to handle grammar in a foreign language. No grammatical terminology is used, so scores do not depend on specific memory for grammatical terms.

[5] Although Appendix A lists possible Index Scores below 20, current scoring devices do not yield Index Scores below 20.

[6] The MLAT was standardized in part on an FSI sample. Although that sample, as a result of the times (late 1950s) was all male, no gender differences have appeared on the MLAT among present students on any subtest of the MLAT or on its Total or standardized score.

Part V. Paired Associates. The examinee is presented with 24 foreign words with their English equivalents and given some time to learn them. The words are then tested. This subtest is said to measure the examinee's ability to memorize by rote--a useful skill in learning new vocabulary in a foreign language.

Raw Score Total. Total of all five subscales.

Index Score. Originally a scaled (T) score used at FSI that is based on the Total. The original mean was 50, with a standard deviation of 10. These norms are now out of date; the Index is now simply a conversion of the raw Total into a scale ranging between 20 and 80. Local norms using the Index have not been formally established because the Index score using the original norms is deeply embedded in the agency's personnel system.

End-of-training proficiency tests. These tests provide the main criterion measure in this study. At the end of training, FSI students are given proficiency assessments resulting in ratings ranging from 0 to 5 for speaking (the S-score, which includes interactive listening comprehension) and for reading (the R-score). The full oral interview, including speaking, interactive listening, and an interactive reading test using authentic materials, takes two hours. R-3, for example, indicates reading proficiency level 3 ("professional" proficiency); S-2 represents speaking proficiency level 2 (working proficiency). Other levels are 0 (no proficiency), 1 (survival level), 4 (full professional proficiency, with few if any limitations on the person's ability to function in the language and culture), and 5 (equivalent to an educated native speaker).

The ratings are equivalent to the guidelines of the Interagency Language Roundtable/American Council on the Teaching of Foreign Languages (ILR/ACTFL) that originated at FSI and have been developed over the years by government agencies. (These guidelines are detailed by Omaggio, 1986). Most students enter FSI with goals of end-training proficiency ratings at S-3 R-3 for full-time training, comparable to ILR/ACTFL Advanced Proficiency.

Reliability studies have shown that government agencies have high interrupter reliability for proficiency ratings within a given agency, but that the standards are not always the same at every agency; thus raters at different government agencies do not have as high an interrater reliability as raters at the same agency. Proficiency ratings are thus considered reliable indicators of the level of language performance of an individual student within an agency (Clark, 1986. "Plus" scores (e.g., indicating proficiency between S-2 and S-3) were coded as 0.5; thus, for example, a score of S-2+ was coded 2.5.

Learning style, strategy, and personality instruments. The Learning Style Profile is a pure learning style instrument. The Myers-Briggs Type Indicator and its Type Differentiation Indicator scoring system is both a personality instrument and a way to assess learning style, as is the Hartmann Boundary Questionnaire. The student learning activities questionnaires tap learning strategies.

The Hartmann Boundary Questionnaire (HBQ) (Hartmann, 1991). The HBQ was developed for research with sleep disorders and nightmares, using a psychoanalytic theoretical base. It is intended to examine the degree to which individuals separate aspects of their mental, interpersonal, and external experience through "thick" or "thin" psychological boundaries. Its 146 items address the following dimensions: sleep/dreams/ wakefulness, unusual experiences,

boundaries among thoughts/feelings/moods, impressions of childhood/adolescence/adulthood, interpersonal distance/openness/ closeness, physical and emotional sensitivity, preference for neatness, preference for clear lines, opinions about children/adolescents/adults, opinions about lines of authority, opinions about boundaries among groups/peoples/nations, opinions about abstract concepts, plus a total score for all twelve of the above scales. Hartmann found women and younger people to score consistently "thinner" than men and older people. Cronbach alpha reliability for the HBQ is .93, and theta reliabilities for subscales are .57–.92 (Hartmann, 1991).

The National Association of Secondary Schools Principals' Learning Style Profile (LSP), (Keefe & Monk, with Letteri, Languis, & Dunn, 1989). This is a 125-item composite measure composed of many different approaches to measuring learning style. The main subscales are cognitive skills (analytic, spatial, categorization, sequential processing, detail memory, discrimination), perceptual response (i.e., sensory preferences: visual, auditory, emotive/kinesthetic), orientations (persistence, verbal risk-taking, manipulative), study time preferences (early morning, late morning, afternoon, evening), and environmental context for learning (verbal vs. spatial, posture, light, temperature, mobility, and grouping). Cronbach's alpha for the subscales ranged from .47 to .76, with an average of .61. Test-retest reliabilities were .36 to .82 after 10 days and somewhat lower after 30 days. Concurrent validity of the LSP's analytic subscale with the Group Embedded Figures Test was .39. Concurrent validity of the perceptual response subscales of the LSP with the Edmonds Learning Style Identification Exercise was .51 - .64. Many of the environmental context subscales of the LSP correlated with Dunn and Dunn's Learning Style Inventory, .23 - .71. All concurrent validity scores are reported in the manual with a significance value < .002.

The Myers-Briggs Type Indicator (MBTI), (Myers & McCaulley, 1985) Form G. This instrument is a 126-item, forced-choice, normative, self-report questionnaire designed to reveal basic personality preferences on four scales: extraversion-introversion (whether the person obtains energy externally or internally), sensing-intuition (whether the person is concrete/sequential or abstract/random); thinking-feeling (whether the person makes decisions based on objective logic or subjective values); and judging-perceiving (whether the person needs rapid closure or prefers a flexible life). Internal consistency split-half reliabilities average .87, and test-retest reliabilities are .70 - .85 (Myers & McCaulley, 1985). Concurrent validity is documented with personality, vocational preference, educational style, and management style (.40 - .77). Construct validity is supported by many studies of occupational preferences and creativity.

The Type Differentiation Indicator (TDI) (Saunders, 1989). The TDI is a scoring system for a longer and more intricate 290-item form (MBTI Form J) that provides data on the following subscales for each of the four MBTI dimensions: extraversion-introversion (gregarious-intimate, enthusiastic-quiet, initiator-receptor, expressive-contained, auditory-visual); sensing-intuition (concrete-abstract, realistic-imaginative, pragmatic-intellectual, experiential-theoretical, traditional-original); thinking-feeling (critical-accepting, tough-tender, questioning-accommodating, reasonable-compassionate, logical-affective); and judging-perceiving (stress avoider-polyactive, systematic-casual, scheduled-spontaneous, planful-open-ended, methodical-emergent). The TDI includes seven additional scales indicating a sense of overall comfort and confidence versus discomfort and anxiety (guarded-optimistic, defiant-compliant, carefree-worried, decisive-ambivalent, intrepid-inhibited, leader-follower, proactive-distractible), plus a

composite of these called "strain."7 Each of these comfort-discomfort subscales also loads on one of the four type dimensions, e.g., proactive-distractible is also a judging-perceiving subscale. There are also scales for type-scale consistency and comfort-scale consistency. Reliability of 23 of the 27 TDI subscales is greater than .50, an acceptable result given the brevity of the subscales (Saunders, 1989).

Student Learning Activities Questionnaires. At the end of training, each student in the study was asked to complete two questionnaires: "CLASSACT" (Ehrman & Jackson, 1992) on relative usefulness of a fairly detailed list of classroom activities (Likert scaled 1-3) and "SELFACT" (Hart-Gonzalez and Ehrman, 1992) on relative usefulness (1-3) of their own study activities and estimated time per week devoted to each. These questionnaires are used here for the first time. Because completion at end of training was voluntary and students were very busy with preparations for departure, the return rate was low, and *N*'s for a number of the items are not adequate for analysis. This and other studies using these two questionnaires are part of their validation. When there are sufficient cases, they will be subjected to reliability analysis and factor analysis.

Data Collection and Analysis

Data collection took place over a two-year period, between 1992 and 1994. Students at each of the two annual major intakes were asked to participate but could decline the invitation; under 5% of the students who were approached chose not to participate. During the 1992-1993 academic year, all French and Spanish students (who start 10 times a year) were also invited to join the study, with the same drop-out rate.

All questionnaires except the MLAT were administered within the first week of training. If a student already had an MLAT record, he or she could arrange for those scores to be included in the research data set; otherwise, MLAT administration took place within the first month of the beginning of training. In this sample, almost all (95%) of the MLAT scores were current, i.e., within the previous 3 years. Proficiency tests were administered at the end of training, after (in most cases) 24 or 44 weeks.

Data analysis in this study on SPSS for Windows 5.0.1 (Norusis, 1992) used correlations, one-way analysis of variance (ANOVA), t-tests, and multiple regression. Correlations of the MLAT were done with end-of-training ratings for speaking and reading proficiency (the FSI proficiency test is described above, under "Instrumentation") and with individual difference variables (see above for listing and descriptions of the instruments). The data used for the correlations between end-of-training proficiency and the MLAT Index for all language categories combined were filtered to equalize expected length of training and proficiency outcomes (to make results of a language like French comparable to those of a language like Chinese).

RESULTS

Distributions

Table 1 shows that the Index Score is somewhat higher for category 2, 3, and 4 languages than for category 1 languages in central tendency and range (see "Sample" for definitions of these categories). The part scores follow the same pattern.

Table 1: MLAT Descriptive Statistics for Index Score

Category	*N*	Mean	SD	Range	Mode	Skewedness	Kurtosis
All Students	343	63	10	21 - 80	70	-.973	1.392
Category 1	169	59	12	21 - 80	61, 70	-.808	.625
Categories 2-3	120	66	8	45 - 80	70	-.462	-.171
Category 4	54	63	10	26 - 78	64	-.900	.770

Minimum possible Index: 20; maximum possible Index: 80.
Category 1: Western European languages; Category 2: Swahili, Indonesian, Malay; Category 3: Eastern European and non-Western languages (except Category 4 languages); Category 4: Arabic, Chinese, Japanese, Korean.

The distributions, with their high central tendencies and reduced space below the ceiling for FSI students, reflect several forms of preselection. The first is that many students have self-selected for foreign affairs careers. Most of these went through their agency's selection process. This process has already probably eliminated some of the students least likely to score well on the MLAT. Second, the MLAT Index Score is used for selection of students in the FSI's parent agency's personnel system, along with other evidence of likely learning, especially evidence of previous language learning success. (Such selection is authorized in the personnel regulations for U.S. Department of State, though it is clearly stated that evidence of learning success overrides the MLAT.)

Selection is done in the State Department's personnel system especially for non-Western-European languages, for which training to the "professional" proficiency level (S-3 R-3) takes 44-88 weeks. Relatively low MLAT students (Index below 55 for category 3 or 60 for category 4 languages) with no other evidence of success are normally sent to Western European languages by preference, hence this is where we find a relatively large range of tested aptitude.

The effect of preselection using the MLAT for category 3 and 4 languages is to make it very difficult to analyze the MLAT's predictive value for these languages in this sample. On the other hand, in view of the expense entailed by 44-week and 88-week intensive language training, assignments personnel understandably seek every indication of likely success or lack of it, without reference to the needs of the researcher.

Other results are described under two rubrics: findings related to prediction of language learning success and findings related to diagnosis and student counseling. The former are quantitative; the latter are qualitative.

Results related to prediction of language learning success

Correlations Correlation coefficients for MLAT Index, Total, and part scores with S- and R-ratings range in the 40's and 50's for the MLAT when a broad range of scores is available, comparable with coefficients found originally by Carroll (1990). The Index Score tends to show higher correlations with end-of-training proficiency ratings than the part scores or the Total. Correlations for the Index Score are shown in Table 2.

Table 2: Correlations of MLAT Index Score with End-of-Training Proficiency Ratings

	r	S-rating	r	R-rating
All languages:	.44	(N = 343)	.40	(N = 341)
Category 1 languages:	.52	(N = 169)	.55	(N = 168).
Category 2-3 languages	.34	(N = 120)	.35	(N = 120)
Category 4 languages	.47	(N = 54)	.34	(N = 53)

Category 1: Western European languages; Category 2: Swahili, Indonesian, Malay; Category 3: Eastern European and non-Western languages (except Category 4 languages); Category 4: Arabic, Chinese, Japanese, Korean. S-rating: speaking and interactive listening; R-rating: reading.

Correlations are weakest for category 2-3 languages and strongest for category 1 languages, where there is the greatest range and the distribution of MLAT scores closely resembles a normal distribution. For categories 1-3, correlations with reading and speaking are roughly the same. In category 4 languages, they are stronger for speaking than for reading. This difference may be because there is less range in reading scores (they are much lower for beginners than in other languages), or possibly because the MLAT does not address abilities needed for reading languages that use Chinese characters--three out of the four category 4 languages.

T-tests Cut points were established such that the cut was made between a score and all those below it. For example, a cut point of S-2 divides between cases less than S-2 and those equal to or greater than S-2. T-tests were done at each cut point from 1+ to 3+ (there were not enough 4-level scores in the sample for meaningful statistics).[7] P-values range from .0001 to .044; with a few exceptions as indicated, only those at the .0001 level are reported in Table 3.

Table 3. Cut points at which the part discriminates at .0001 significance

MLAT Part	ILR Speaking Level	ILR Reading Level
Part I	S-2, S-3	
Part II	S-2	R-2
Part III	S-2, S-2+, S-3	R-2, R-2+, R-3
Part IV	S-2, S-3	R-2
Part V	S-2	
Total Score	S-1+, S-2, S-3, S-3+[8]	R-2, R-2+, R-3, R=3+
Index Score	S-1+, S-2, S-3, S-3+[8]	R-2, R-2+, R-3, R=3+

[7] A table of the T-test results is available on request.

[8] The Total Score discriminates at the S-3+ cut point at a significance of .012, and the Index Score at the .013 level.

The best discriminators at all levels of proficiency appear to be Parts III (English vocabulary in altered spellings) and the Total and Index Scores. Another summary of the same results appears in Table 4, this time organized by cut point:

Table 4. Part(s) discriminating at .0001 significance at each ILR score cutpoint

Speaking:	S-1+	Total Score
	S-2	Parts I, II, III, IV, V Total Score, and Index Score
	S-2+	Part III, Total Score, and Index Score
	S-3	Parts I, III and IV, Total Score, and the Index Score
	3+	The Total and the Index Scores[8]
Reading:	R-1+	None
	R-2	Parts II ,III, and the Total Score
	R-2+	Part III and the Total Score.
	R-3	Part III, the Total Score, and Index Score
	R-3+	The Total Score and the Index Score.

Analysis of Variance: This investigation was done only for the entire sample, because the numbers of subjects were not sufficient for category 2–3 or 4 languages separately. In a study of the extremely strong and weak students in the sample, the bottom 3–4 percent were contrasted against all others and top 5–6 percent against all others. Extreme students were selected on a formula that combined length of training, relative difficulty of language by category, and end-of-training scores. There were fewer students at the low end because the very weakest may be withdrawn well before scheduled end of training and because both teachers and students make every effort to reach the student's training goal, which in most cases is S-3 R-3. More detail on the extremes study, including the selection formula, is available in Ehrman (1994b).

Data for the individual difference variables were analyzed using the one-way analysis of variance procedure in SPSS for Windows 6.1. The findings for the MLAT are displayed in Tables 5 and 6.

Speaking: Of all the variables analyzed, the Parts III, IV, V, the Total, and the Index scores best differentiated the *weakest* students, that is, these variables had the largest F-scores. The MLAT variables also differentiated these weak students better than any other of the many variables in the study.

For the *strongest* students' speaking scores, the Index ($F=7.83$, $p < .0055$) was the strongest differentiator from among the MLAT and learning style variables, but it was not as good as these biographical background variables: education level, number of previous languages, and previous highest score in speaking and especially reading. The MLAT appears to differentiate the strongest speakers less clearly than the weakest speakers and readers and the strongest readers.

Table 5: Speaking Performance Extremes: ANOVAs

Weakest, Speaking	*N* selected (weakest): 4 (Parts & Total), 6 (Index) *N* not selected (all others)= 292 (Parts & Total), 337 (Index).					
Part	**Weakest Mean**	**All Others Mean**	**Weakest SD**	**All Others SD**	**F**	**sig.**
I	24.5	36.5	6.5	9.1	6.8524	.0093
II	18.5	24.7	3.5	4.5	7.3634	.0070
III	11.0	28.3	8.6	9.9	12.1415	.0006
IV	15.3	28.0	5.3	7.5	11.4289	.0008
V	11.5	19.3	4.7	5.3	11.4289	.0008
Total	80.8	136.7	24.6	27.5	16.3881	.0001
Index	43.2	62.7	10.8	10.5	20.5548	.0000

Strongest, Speaking	*N* selected (strongest): 14 (Parts & Total), 19 (Index) *N* not selected (all others) = 281 (Parts & Total), 324 (Index).					
Part	**Weakest Mean**	**All Others Mean**	**Weakest SD**	**All Others SD**	**F**	**sig.**
I	40.5	35.0	4.9	9.7	4.4395	.0362
II	27.1	24.3	2.8	4.7	5.2765	.0225
III	32.8	27.0	7.0	14.2	4.5701	.0336
IV	30.0	27.2	5.0	7.9	1.7067	.1927 ns
V	20.8	18.8	4.2	5.5	1.6950	.1942 ns
Total	151.2	132.5	13.8	29.6	5.7291	.0175
Index	68.2	60.9	5.9	11.2	7.8286	.0055

Data analysis done by SPSS for Windows v. 6.1, One Way Analysis of Variance Test. Degrees of freedom are available on request.

Reading For reading, Parts III and IV and the Total and Index Scores best differentiate the *weakest* students. The *strongest* are differentiated clearly by all MLAT parts except Part IV; with the Index Score providing the clearest distinction.

Table 6: Reading Performance Extremes: ANOVAs

Weakest, Reading	*N* selected (weakest): 3 (Parts & Total), 4 (Index) *N* not selected (all others) = 292 (Parts & Total), 337 (Index).					
Part	**Weakest Mean**	**All Others Mean**	**Weakest SD**	**All Others SD**	**F**	**sig.**
I	23.0	36.4	7.0	9.1	6.4559	.0115
II	17.7	24.7	3.8	4.5	7.1481	.0079
III	7.3	28.2	5.5	9.9	13.4109	.0003
IV	13.0	28.0	3.5	7.5	11.8901	.0006
V	11.0	19.3	5.6	5.3	7.3757	.0070
Total	72.0	136.6	21.2	27.6	16.3758	.0001
Index	40.5	62.7	12.6	10.5	17.6391	.0000

Strongest, Reading		N selected (strongest): 78 (Parts & Total), 93 (Index) N not selected (all others)= 217 (Parts & Total), 248 (Index)				
Part	Weakest Mean	All Others Mean	Weakest SD	All Others SD	F	sig.
I	38.9	33.8	6.3	10.5	15.0647	.0001
II	26.1	23.7	3.5	4.8	15.4653	.0001
III	31.0	26.9	8.6	10.2	14.7692	.0002
IV	29.2	26.7	6.5	7.9	6.1293	.0140
V	21.3	17.9	4.1	5.6	22.5703	.0000
Total	146.5	128.0	20.9	30.1	23.7211	.0000
Index	66.3	59.6	8.0	11.3	26.1914	.0000

Data analysis done by SPSS for Windows v. 6.1. One Way Analysis of Variance Test. Degrees of freedom are available on request.

Multiple Regression: Multiple regression analysis for end-of-training speaking and reading examined the effects of age, education level, number of previous languages studied, highest previous speaking and reading ratings, a general motivation rating, two self-efficacy ratings (self-rated aptitude and expectation of success in this course), two anxiety ratings (for the course in general and about speaking in class), and the MLAT Index Score.

For speaking, the analysis yielded a multiple R of .40, R Square of .16, with two predictors in the equation: the MLAT Index Score (Beta .32, T = 3.293 p = .0014) and Highest Previous Reading Score (Beta .21, T = 2.208, p = .0297).

For reading, the analysis yielded a multiple R of .37, R Square of .14, with the same two predictors in the equation: the MLAT Index Score (Beta .27, T = 2.798, p = .0063) and Highest Previous Reading Score (Beta .22, T = 2.266, p = .0258).

Results related to diagnosis and student counseling

In this section, both quantitative and qualitative findings are described, as part of an ongoing effort to build learner profiles that can be used by teachers, teacher trainers, program managers, and even students themselves to enhance student learning. The quantitative results contribute to a fuller picture of the kinds of students who are advantaged and disadvantaged in full-time intensive and largely communicative language training, by adding personality factors to more cognitive abilities. The qualitative material is very exploratory, but it has been promising enough to merit description here so that others can use and test the emerging patterns. It is also included here because it provides more information on what the MLAT may actually be measuring, and because it sheds more light on the complexity of the apparently simple factor-analysis-based MLAT parts.

Relationships with Other Individual Difference Variables: There are other variables than the MLAT that are useful in the building of an individual learner profile that can be used for diagnosis and counseling (the utility of these for prediction is more directly addressed in Ehrman, 1993, 1994a, b, 1995, 1996, Ehrman & Oxford, 1995; Oxford & Ehrman, 1995). These variables bear interesting relationships to the MLAT. Correlations of at least .30 between the MLAT Index

Score and/or Total Score and other instruments used in the larger study are presented in Table 7. The correlations suggest the relationships described below.

Table 7: MLAT Index or Total Score Correlations with Other Variables

Variable	Lang. Category Grp	rho	Correlate	N
Number of Previous Langs	All	.40**	Index	245
HBQ Prefer Blurred Edges	Cat. 1	.51*	Total	25
HBQ Prefer Low Neatness	Cat. 1	.47	Total	25
HBQ Thin External Boundaries	All	.32**	Total	102
HBQ Total Score (thin)	All	.30**	Index	110
MBTI/TDI Intellectual (N)	Cat. 1	.45**	Index	96
MBTI/TDI Intellectual (N)	Cat. 2-3	.35**	Index	103
MBTI Intuition	Cat. 1	.34**	Total	93
MBTI Imaginative (N)	Cat. 1	.34**	Index	96
MBTI Introversion	Cat. 1	.30*	Total	93
LSP Simultaneous Processing	Cat. 1	.45	Index	24
LSP Sequential Processing	Cat. 1	.43	Index	24

*All the above correlations are significant at least at the .05 level; * indicates the .01 level; ** indicates the .001 level. HBQ: Hartmann Boundary Questionnaire. MBTI: Myers-Briggs Type Indicator. LSP: Learning Style Profile. "Imaginative" and "Intellectual" represent the intuition (N) poles of the MBTI TDI Realistic-Imaginative and Pragmatic-Intellectual subscales for the sensing-intuition main scale.*

Those who have scored high on the MLAT tend to have studied languages previously, often prefer an "intuitive" approach to taking in information on the MBTI. MBTI intuition indicates preferences for the abstract over the concrete, search for meaning, a preference for the "big picture" rather than details, and the speculative over the strictly experiential (Myers & McCaulley, 1985). They describe themselves as having relatively thin ego boundaries, especially with respect to such matters as dislike for too much neatness, order, and clear-cut separations among visual images. Thin ego boundaries, correlated with MBTI intuition, indicate receptivity to a wide range of experience, both internal and external, and a willingness to blur categories. This concept is used to operationalize a model of tolerance of ambiguity (Ehrman, 1993). High-MLAT students also are often more skilled at simultaneous and sequential visual processing on the LSP.

The analyses of variance in the extremes study support these findings for extremely strong and weak students and add as an advantage a preference for a flexible approach shown in the perceiving pole of one of the MBTI/TDI JP subscales, methodical vs. emergent. (This subscale of the TDI scoring of the long MBTI opposes a desire to know in advance what will happen in contrast with a preference to let events "emerge" and cope with them as they come up; the strongest students indicated a preference for an emergent approach.)

The MLAT and Learning Activities. A recent correlation study showed interesting relationships between the MLAT and a set of activities that students rated for perceived utility both before starting training and at the end of training (Ehrman, 1995). The correlations were

similar for both pre- and post-testing. Though the correlations were generally low (mostly 20's and some in the 30's), there seemed to be suggestive patterns in them when subjected to a content analysis. Findings described below were based on the content analysis of those items with which the MLAT was correlated and on correlations of MLAT scales with variables from the other instruments (Table 7).

In summary, high MLAT Index and *all* part scores correlate with items that are interpreted as reflecting self-confidence as a language learner and tolerance of ambiguity (low-structure activities and input).

The Index and parts II, III, IV, and V are correlated with items suggesting acceptance of/preference for use of authentic material for reading and listening and authentic conversation.

Parts III and IV are correlated with items suggesting endorsement of learning activities that reflect an analytic, structured approach. This effect was slightly stronger for part III; students who rejected a "touchy feely" approach on one item (the only such item) also tended to be high scorers on part III.

In contrast, a more experiential, kinesthetic approach may be suggested by the Index and a peak on part II, at least as indicated by the correlations with preferred learning activities.

Students who endorsed activities that were interpreted as indicating a preference for directing their own study tended to do well on the Index and parts II and IV.

Interpreting part-score profiles. The above patterns suggested possible uses for the MLAT profile in student counseling, where they currently being tested. Some profiles that these data suggest are outlined in below.

1. All parts high (a very high Index will usually represent this kind of profile):
   - has done well on all the parts
   - self-confident as learners
   - respond well to activities that require tolerance of ambiguity
   - like relatively unstructured learning
   - enjoy and even prefer authentic input

A related analysis found a relationship between endorsement of relatively unstructured, ambiguous, authentic activities and higher end-of-training scores (Ehrman, 1995).

2. A more uneven profile in which parts III (especially) and IV are high:
   - analytic learner, perhaps field independent
   - likes a program with a clear plan (<u>not</u> the same as a restrictively sequential program).
   - usually has good knowledge of English vocabulary and grammar.
3. An uneven profile in which Part II is highest, together with a strong Index, (most other parts above average) may indicate a student who likes experiential, hands-on, participatory learning.

4. An uneven profile in which Parts II and IV relatively high, together with a strong Index, may suggest a student who likes to take control of his or her own learning sequence and can use both analytic and global learning strategies comfortably.
5. When either part I or part V is the highest of the part scores, there so far seems to be little that is distinctive, though interviews are suggesting that low scores on part V appear to indicate either poor mnemonic skills or weak metacognitive strategies, or both.
6. All parts low (a very low Index will usually represent this kind of profile):
   has done poorly on all the parts
   often lacks self-confidence as a learner and subject to anxiety because of slow progress
   likely to be overwhelmed by unstructured and uncontrolled input
   will need a great deal of scaffolding for longer than most other students
   likely to progress slowly

Overall Total score on the MLAT or the Index gives a crude measure useful when it is either very low or very high: a very low Total or Index score indicates weakness in all the factors; a very high score suggests strength in all the factors. When the Index falls in the middle range--roughly within a standard deviation of the mean-- it becomes much more important to examine the "scatter" of the part scores.

Using part scores with students. The student counseling activity uses the variations in part scores to initiate interpretations that are raised with the student to examine how he or she learns. Interpretation usually requires an interview of the student. Responses by students to the question "What happened when you were doing this part?" provides useful information about the skills tested by each part. Each of the MLAT factors probably represents a set of abilities. For example, Part III has proved particularly fruitful in the diagnostic process with students. Among the possible task requirements of this item are: gestalt processing of the whole word; sound-symbol processing; rapid hypothesis testing of sound-symbol possibilities; shift in mental set; and semantic evaluation.

These task requirement possibilities are represented as student performance in the following six cases of poor outcome on Part III, each of which is followed by implications for the classroom. The cases represent composites of responses actually received to the query about what happened while students were completing this subtest. (Many examples of real cases with specific score profiles, are to be found in Ehrman, 1996.)

1) One student might have done poorly on Part III because of difficulty with the kinds of analytic activities often described as "field independent." This student is likely to have difficulty with induction of rules and patterns and with grammar-oriented activities that have little context. Students of this sort usually find more contextual learning helpful.

2) Another might do poorly on the same part because of a weak English vocabulary (among the possible causal factors: poor education, low intelligence). This student, if a native speaker of

English,[9] may have difficulty with vocabulary learning (among other things) because of lacking concepts and background. The classroom may have to include activities to help this student build content background as well as language.

3) A third experiences difficulties reorganizing schemata or with gestalt processing or shifting mental set. Part III makes considerable demands on a person's ability to shift mental set. Such a student may be more comfortable with relatively predictable activities and less so with open-ended ones and may need assistance in building skills for coping with the unfamiliar or unexpected.

4) Yet another might have a *phonetic coding difficulty* of the sort described by Sparks, Ganschow et al. (1991), i.e., working with sound-symbol relationships. He or she is likely to have corresponding low scores in Parts I and II, which also require decoding of sounds. Such a student is likely to be handicapped in both speaking and reading and will need more time to absorb material. Kinesthetic input is likely to help this student.

5) Links among extraversion, desire for language use outside the classroom, and MLAT Part III suggest a *distractibility* factor. That is, a strongly extraverted student who is drawn to interpersonal interactions might not be as adept at the kind of focus that the puzzle solving aspect of Part III entails as one who tunes out the world more readily. Study strategies, including frequent breaks and setting up conditions to maximize concentration, might help a student who has difficulty concentrating.

6) Finally, a person who is reminded by Part III items of crossword puzzles and dislikes them has had an *affective* reaction which interferes with ability to use cognitive resources. Alternatives to "puzzle-solving" activities would probably help the sixth student, or perhaps cooperative learning when puzzle-like activities are part of the curriculum. The teacher would need to be alert to the affective impact of these activities.

Interpretation of a student's profile is made more complex by factors that can affect any or all of the parts of the test. In some cases, a low score on Part III (or any other part) may be the result of a mechanical error, such as marking in the wrong row of the answer sheet. sometimes a student will say that he or she did not understand the instructions for a given part. (This response raises questions about attention, motivation, or test-taking strategies.) Some students ascribe low scores to fatigue, which is plausible especially for the later parts. Interpretation is further complicated by the fact that a student might suffer from several of these difficulties at once.

DISCUSSION

Summary. Despite the effects of restricted range, skewed distribution, and relatively limited ceiling (because of negative skew for this high-end sample), the MLAT remains the best predictor of the variables examined. In general, the Index Score is the most useful of the MLAT variables

[9] The MLAT is designed for use with native speakers of English. At FSI it is considered invalid for non-native speakers, though if one takes it and does well (Index greater than 50), such performance is considered a promising sign. Low scores, on the other hand, are ignored.

as a predictor (strong in all cases, and with highest correlation coefficients). Of the part scores, Part III is the strongest predictor. Part III, with its dependence on knowledge of English vocabulary as well as ability to solve puzzles, may also be an indirect indicator of general intelligence. This would apply to both fluid ability, because of the cognitive restructuring required by the task, and crystallized ability (vocabulary), and "g" or general intelligence, since general vocabulary is also considered to be the single best stand-in for overall intelligence (Anastasi, 1988, Wesche, Edwards, and Wells, 1982).

Is the MLAT more suitable for Western European languages than for non-Western languages? The question remains open. Correlations and T-tests show stronger results for category 1 languages than for 2, 3, and 4 languages. On the other hand, the substantial preselection of students suggested by the very skewed distribution and the restriction of range in the sample may account for this finding as much as appropriateness of the MLAT for non-European languages. Furthermore, the fact that the correlations for category 4 language outcomes are actually better than those for category 3 languages, despite substantial truncation of range, might suggest that the MLAT is actually a fairly strong predictor for these languages. (The higher correlations might also be related to the much smaller *N* for category 4 languages.) We cannot test either hypothesis on the FSI language-student population as long as they are preselected and preselected using the MLAT.

Of the extended set of variables in the research project (including learning strategies, cognitive styles, motivation, anxiety, and personality variables), the MLAT Index Score also continues to be the strongest, both in the correlation coefficients and ANOVAs of extremely weak and strong students. It is especially powerful as a selector of extremes.

In addition to the relatively crude information provided by the Index score that may help in selection for training, the part-score profile shows promise as a way to better target classroom interventions and advice to students about appropriate learning strategies to develop. High performance on the MLAT appears to be related to personality variables that indicate high tolerance for ambiguity and the ability to reconceptualize input.

Is the MLAT passé in an age of communicative teaching? The MLAT has been criticized by many as rating aptitude only for audio-lingual training, which was in vogue when the MLAT was developed. However, the MLAT correlations remain about the same, even though the teaching methodology has changed considerably (most FSI courses now have a substantial communicative component, and some are almost wholly communicative). Why is this so? The following are some possibilities.

1. Perhaps the MLAT is really multidimensional, and a different set of dimensions applies to different methodology.
2. Perhaps the operative factor is really some form of coping with ambiguity or coping with the unfamiliar.
3. Possibly it is the "g" (general intelligence)-factor that is operative for FSI students. (Sasaki (1993) found a general found a general cognition factor, which she describes as similar to

"g" to account for 42% of the variance among Japanese college students studying English as a foreign language.)

4. The very nature of classroom training may make a difference. Although FSI classroom training requires the ability to cope with communicative activities and access global and inferential learning, it also makes heavy demands on analytic skills. These may become increasingly important at higher proficiency levels; this fact may be why part III, which is most strongly associated with analytic learning, differentiates most at the higher levels in the T-tests and why parts III and IV together are the most predictive of extremes in achievement, together with the Index, which is more associated with predilection for the more open-ended learning that is also necessary for achieving high proficiency levels in FSI classrooms. The study of ego boundaries using the Hartmann Boundary Questionnaire (Ehrman, 1993) found a similar construct, labeled "tolerance of ambiguity" to be essential to effective classroom learning at FSI. In this study, thin ego boundaries that let a student take in new data were not enough alone--students had to impose some sort of mental structure on their intake and at the same time stay open to the fact that their structures were hypothetical. Investigation now under way is examining the applicability of the field independence construct to these findings, further information on which is to be found in Ehrman, 1996.

The aptitude concept. Expanding the aptitude concept is one of the subjects of an ongoing investigation of individual differences in language learning. The subject is discussed in greater detail in Ehrman, 1994b, 1995, 1996

Among the outcomes of the study is evidence for an expanded definition of aptitude that includes both cognitive aptitude (measured specifically for languages by the MLAT and more generally by cognitive aptitude tests?) and personality factors that predispose a learner to cope with ambiguity and apparent chaos. These become especially important in the relatively unstructured learning setting of communicative teaching approaches. A nexus is emerging of the following characteristics that seem to be related to success in the demanding intensive FSI classroom::

- cognitive aptitude (may include ability to cope with the unfamiliar)
- random (vs. sequential) learning
- orientation to meaning over form
- ability to cope with surprises (linguistic and pedagogical)
- openness to input and tolerance of ambiguity
- ability to sort input, analyze as appropriate, and organize into mental structures.

The last is almost certainly related to the field independence construct in some way; it may be that the MLAT provides a way to measure field independence through verbal activities, in contrast to the usual tests of ability to disembed geometric figures. Such a measure might improve the value of the field independence construct for language learning.

Absence of the above-listed characteristics appears to disadvantage FSI learners, perhaps more than the presence of these variables advantages those learners (Ehrman, 1994a, b, 1995, 1996).

There seems to be a kind of aptitude-personality nexus that consists of cognitive flexibility, tolerance of ambiguity (including ability to impose structure on input), and ability to make use of random access strategies.

The MLAT is the most powerful of the predictive variables used, even in programs that are very different from those in vogue when it was designed. It may be that the ability to manage unfamiliar and contradictory input leads both to success in communicative classrooms and to high scores on the MLAT. The MLAT may gain its relative power because it requires the examinee to cope with the unfamiliar on tasks that at least partially simulate language learning tasks, whereas personality inventories are asking about general life preferences, and strategy inventories do not address how the strategies are used but only whether the student is aware of using them. "Faking good" is nearly impossible on the MLAT, and malingering is vanishingly rare at FSI.

Although the MLAT provides strong information about classroom language learning ability, it is supplemented by personality variables. The significant correlations between the MLAT and the personality measures, though not strong (between .21 and .33), are consistent across personality questionnaire and MLAT subscales (Ehrman 1993, 1994a, b, 1995). In all cases, MLAT scores are linked with variables that suggest tolerance for ambiguity.[10]

The links between the MLAT and personality variables suggest a role for the disposition to use one's cognitive resources in ways that go beneath the surface and that establish elaborated knowledge structures. Those who are open to new material, can tolerate contradictions, establish hypotheses to be tested, focus on meaning, and find ways to link the new with previous knowledge structures seem to have an advantage in managing the complex demands of language and culture learning. The weakest students appear to be overwhelmed by the chaos they encounter; the strongest meet it head on, may even embrace it to a degree.

As of now, the answer to the question "is the MLAT passé?" is: probably not, though it has much the same limitations as a sole *predictor* of learning success that it has always had. It is pretty good, especially if viewed as an indicator of learning dispositions that will affect classroom performance, but it probably should not be more than one tool in a toolkit. Scatter analysis of the part scores is a promising use for placement, counseling, and remediation, particularly in the hands of an evaluator who treats the scores as signposts to interpretations to be tested, not as absolute predictors.

Limitations of this study: The greatest limitation of this study, like all those from FSI, is the question of generalizability. Use of a sample drawn from a high-end, preselected population in itself restricts range, affects distributions, and strongly indicates the need for replication with samples more typical of what the usual reader of this publication works with. For the MLAT,

---

[10] A very recent study also shows a correlation of the MLAT with self-report of 'field sensitivity' (Index, r=.58, Part II .61, Part III .46, all at a p level of 0001). Field sensitivity, discussed at greater length in Ehrman, 1996, in press, and Ehrman & Leaver, 1997, can be defined as preference for working with new material in context, in stories or articles or at least sentences. Field sensitive learners often pick up new words, ideas, etc. peripherally, without planning in advance; they can be described as using of a floodlight to learn in contrast to field independence, which uses a spotlight.

unlike any of the other instruments in the larger study, the use of the instrument itself to help preselect the sample severely limits both the statistical normality of the sample and our ability to make inferences from the findings.

The impossibility of establishing a truly normal distribution of MLAT scores in this sample also means that the statistical tests that assume normal distributions and similar sample sizes are used in unconventional ways. The number of tests conducted increases the chance of type I errors (false positives), though the consistency of findings over a number of variables may reduce the likelihood of such error. For these reasons, the findings reported here must be considered suggestive, not conclusive.

The qualitative investigation has been undertaken on an ad hoc basis and therefore consists for now of working hypotheses about the meanings of high and low points in MLAT part-score profiles. It has yet to be investigated more systematically at a level beyond individual cases.

Next Steps. There is much more to look at in these data, in the course of trying to find out what the MLAT is good for and what its limitations are. Among these are to seek normally distributed samples on which to replicate this study, begin multiple regression and discriminant analysis to see if MLAT is a better predictor in combination with other variables; and to find out what happened with subjects who return from overseas and are tested--did they improve, get worse, stay the same? On the qualitative front, continued investigation can seek to confirm the working hypotheses described above in the section on student counseling and systematize them for use by people other than researchers, so that the MLAT part scores can provide useful information about specific learning strengths and difficulties that can be used in curriculum design and interventions with individual students. Eventually, a quantitative study of the part-score profiles should be designed and undertaken.

References

Adams, M. (1984). The Modern Language Aptitude Test (Percentile Ranks). Unpublished FSI document. Arlington, VA: Foreign Service Institute.

Anastasi, A. (1988). *Psychological testing*. 6th ed. New York: Macmillan.

Campbell, C. (1987). Survey of Attitudes Specific to the Foreign Language Classroom. Unpublished manuscript

Carroll, J. (1962). The prediction of success in intensive foreign language training. In Glaser, R. (Ed.), *Training research and education*. Pittsburgh, PA: Univ. of Pittsburgh.

Carroll, J. (1990). Cognitive abilities and foreign language aptitude: Then and now. In Parry, T. & C. W. Stansfield (Eds.), *Language aptitude reconsidered*. (pp. 11-29). Englewood Cliffs, NJ: Prentice Hall.

Carroll, J. & Sapon, S.M. (1959). *Modern Language Aptitude Test*. New York: Psychological Corporation.

Clark, J. (1986). *A study of the comparability of speaking proficiency across three government language training agencies*. Washington, DC: Center for Applied Linguistics.

Demuth, K.A., & Smith, N.B. (1987). The foreign language requirement: An alternative program. *Foreign Language Annals, 20*, 67-77.

Ehrman, M.E. (1993). Ego boundaries revisited: Toward a model of personality and learning. In J.E. Alatis (Ed.). *Strategic interaction and language acquisition: Theory, practice, and research*. Washington, DC: Georgetown University Press.

Ehrman, M.E. (1994a). The Type Differentiation Indicator and adult language learning success. *Journal of Psychological Type, 30*, 10 - 29.

Ehrman, M.E. (1994b). Weakest and strongest learners in intensive language training: A study of extremes. In C. Klee (Ed.) *Faces in a crowd: Individual learners in multisection programs*. Boston MA: Heinle & Heinle.

Ehrman, M.E. (1995). Correlations between MLAT scales and preferred student learning activities. Unpublished data.

Ehrman, M.E. (1995). Personality, language learning aptitude, and program structure. Alatis, J. (Ed.), Linguistics and the education of second language teachers: *Ethnolinguistic, psycholinguistic, and sociolinguistic aspects* (pp. 328-345). Washington DC: Georgetown University Press.

Ehrman, M.E. (1996). *Understanding second language learning difficulties*. Thousand Oaks, CA: Sage Publications.

Ehrman, M.E. (in press). Field independence and field sensitivity. In Reid, J. (Ed.), *Centering on the learner*. Englewood Cliffs, NJ: Prentice Hall Regents.

Ehrman, M.E. & Jackson, F.H. (1992). Classroom activities survey. Unpublished manuscript.

Ehrman, M.E. & Oxford, R.L. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal* (1), 67-89.

Ehrman, M.E., & Leaver, B. L. (1997). Sorting our global and analytic functions in second language learning. American Association for Applied Linguistics annual meeting, Orlando, FL, March 1997.

Hart-Gonzalez, L.H. & Ehrman, M.E. (1992). Study Activities Questionnaire. Unpublished manuscript.

Hartmann, E. (1991). *Boundaries in the mind: A new psychology of personality*. New York: Basic Books.

Oxford, R.L., & Ehrman, M.E. (1995). Adults' language learning strategies in an intensive foreign language program in the United States. *System, 23*, 359-386.

Jackson, F.H. (1994). Language aptitude. Draft talking points prepared for FSI roundtable on the Modern Language Aptitude Test, Annual meeting of the Language Testing Research Colloquium, Washington, DC.

Keefe, J.W. & Monk, J.S., (with Letteri, C.A., Languis, M., & Dunn, R). (1989). *Learning Style Profile*. Reston, VA: National Association of Secondary School Principals.

Lefrancois, J. & Sibiga, T. C. (May, 1986). Use of the Modern Language Aptitude Test (MLAT) as a diagnostic tool. Unpublished paper. Source unknown.

Lett, J.A., & O'Mara, F.E., (1990). Predictors of success in an intensive foreign language learning context: Correlates of language learning at the Defense Language Institute Foreign Language Center. In Parry, T. & C. W. Stansfield (Eds.), *Language aptitude reconsidered*. (pp. 222-260.) Englewood Cliffs, NJ: Prentice Hall.

Lett, J.A., & Thain, J. (1994). The Defense Language Aptitude Battery: What is it and how well does it work? Paper delivered at the Language Aptitude Invitational Symposium, Arlington VA.

Myers, I.B., & McCaulley, M.H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.

Norusis, M. J. (1992). *SPSS for Windows 5.0.1*. Chicago: SPSS Inc.

Parry, T.S. & Child, J.R. (1990). Preliminary investigation of the relationship between VORD, MLAT, and language proficiency. In Parry, T. & C. W. Stansfield (Eds.), *Language aptitude reconsidered*. (pp. 30-66.) Englewood Cliffs, NJ: Prentice Hall.

Parry, T., & Stansfield, C.W. (1990). Introduction. In Parry, T. & C. W. Stansfield (Eds.), *Language aptitude reconsidered*. (pp. 1-10.) Englewood Cliffs, NJ: Prentice Hall.

Parry, T., & Stansfield, C. W. (Eds.) (1990). *Language aptitude reconsidered*. Englewood Cliffs, NJ: Prentice Hall.

Petersen, C.R., & Al-Haik, A.R. (1976). The development of the Defense Language Aptitude Battery (DLAB). *Educational and psychological measurement, 6*: 369-380.

Pimsleur, P. (1966). *The Pimsleur language aptitude battery*. New York: Harcourt, Brace, Jovanovich.

Pimsleur, P. (1968). Aptitude testing. *Language Learning*. Special Issue Number 3, August 1968, 73-78.

Sasaki, M. (1993). Relationships among second language proficiency, foreign language aptitude, and intelligence: A structural equation modeling approach. *Language Learning 43*, 313-344.

Saunders, D. (1989). *Type Differentiation Indicator Manual: A scoring system for Form J of the Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.

Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.

Skehan, P. (1991). Individual differences in second language learning. *Studies in Second Language Acquisition 13(2)*, 275-278.

Sparks, R. L., Ganschow, L., and Patton, J. (1995). Prediction of performance in first-year foreign language courses: Connections between native language and foreign language learning. *Journal of Educational Psychology, 87*: 638-655.

Spolsky, B. (1995). *Measured words: The development of objective language testing*. New York: Oxford.

Thain, J. (1992). DLAB II prototype development: Status report and CY92 plan. Technical report. Monterey, CA: Defense Language Institute.

Wesche, M. B. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K.C. Diller (Ed.), *Individual differences and universals in language learning aptitude*. Rowley, MA: Newbury House.

Wesche, M. B., Edwards, H., & Wells, W. (1982). Foreign language aptitude and intelligence. *Applied Psycholinguistics, 3*, 127-40.

Wesche, M., Edwards, H., & Wells, W. (1982). Foreign language aptitude and intelligence. *Applied Psycholinguistics*, 127-140.

Wilds, C. P. (1965). MLAT Index Scores: Foreign Service Institute Curve - 1965. Unpublished table. Arlington, VA: Foreign Service Institute.

Appendix A: Conversion Table for MLAT Raw Total and Index Scores

Raw Total	Index	Raw Total	Index	Raw Total	Index
0-9	15	67-68	37	125-127	59
10-12	16	69-71	38	128-129	60
13-15	17	72-74	39	130-132	61
16-18	18	75-76	40	133-135	62
19-21	19	77-79	41	136-137	63
22-23	20	80-82	42	138-140	64
24-26	21	83-84	43	141-143	65
27-29	22	85-87	44	144-145	66
30-31	23	88-90	45	146-148	67
32-34	24	91-92	46	149-150	68
35-37	25	93-95	47	151-153	69
38-39	26	96-97	48	154-156	70
40-42	27	98-100	49	157-158	71
43-44	28	101-103	50	159-161	72
45-47	29	104-105	51	162-164	73
48-50	30	106-108	52	165-166	74
51-52	31	109-111	53	167-169	75
53-55	32	112-113	54	170-172	76
56-58	33	114-116	55	173-174	77
59-60	34	117-119	56	175-177	78
61-63	35	120-121	57	178-180	79
65-66	36	122-124	58	181-180	80

From Wilds (1965)